Attorney Docket: 81862.P247 *PATENT*

UNITED STATES PATENT APPLICATION

FOR

SYSTEM AND METHOD FOR DISTRIBUTING GUARANTEED BANDWIDTH AMONG SERVICE GROUPS IN A NETWORK NODE

FIRST NAMED INVENTOR:

KENT WENDORF

PREPARED BY:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 WILSHIRE BOULEVARD
SEVENTH FLOOR
LOS ANGELES, CA 90025-1026

(408) 720-8300

"Express Mail" mailing label number EL 672 751 323 US

Date of Deposit 4/19/01

I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231

Juanita Briscoe
(Typed or printed name of person mailing paper or fee)

Juanita Briscoe 4/19/01
(Signature of person mailing paper or fee) *Date*

SYSTEM AND METHOD FOR DISTRIBUTING GUARANTEED BANDWIDTH AMONG SERVICE GROUPS IN A NETWORK NODE

FIELD OF THE INVENTION

[0001] The present invention relates generally to telecommunications systems and, more particularly, to a system and method for distributing guaranteed bandwidth among service groups in a network node.

BACKGROUND OF THE INVENTION

[0002] An Asynchronous Transfer Mode (ATM) is a switching and multiplexing technique designed for transmitting digital information, such as data, video, and voice, at high speed, with low delay, over a telecommunications network. The telecommunications network, for example an ATM network, includes a number of switching nodes coupled through communication links. In the ATM network, bandwidth capacity is allocated to fixed-sized data units named "cells." The communication links transport the cells from a switching node to another. These communication links can support many virtual connections, also named channels, between the switching nodes. The virtual connections assure the flow and delivery of information contained in the cells.

[0003] Each cell contains a cell header and cell data. The cell header includes information necessary to identify the destination of that cell. The components of the cell header include, among other things, a Virtual Channel

Identifier (VCI) and a Virtual Path Identifier (VPI), for collectively identifying an ATM connection for that particular cell, and a Payload Type Identifier (PTI), for indicating whether the cell is sent from one user to another, whether cell data refers to administration or management traffic, and whether congestion is present within the network.

[0004] The ATM Forum, which is a user and vendor group establishing ATM standards, has also defined several ATM class of service categories, used in characterization of a virtual connection, for example, (1) a Constant Bit Rate (CBR), which supports a constant or guaranteed rate to transport services, such as video or voice, as well as circuit emulation, which requires rigorous timing control and performance parameters; (2) a Variable Bit Rate (VBR), real time and non real time, which supports variable bit rate data traffic with average and peak traffic parameters; (3) an Available Bit Rate (ABR), which supports feedback to control the source rate in response to changed characteristics in the network; and (4) an Unspecified Bit Rate (UBR).

SUMMARY OF THE INVENTION

[0005] A system and method for distributing guaranteed bandwidth among service groups in a network node are described. A position on a time scale of a buffer containing multiple data units is determined. A signal prompting selection of said buffer for release of at least one data unit of the data units is modified based on the position determined on the time scale.

[0006] Other features and advantages of the present invention will be apparent from the accompanying drawings and from the detailed description that follows.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The present invention is illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements and in which:

[0008] **Figure 1** is a block diagram of one embodiment for a network.

[0009] **Figure 2** is a block diagram of one embodiment for a network node within the network.

[0010] **Figure 3** is a block diagram of one embodiment for a line card within the network node.

[0011] **Figure 4** is a block diagram of one embodiment for a memory module within the line card.

[0012] **Figure 5** is a flow diagram of one embodiment for a method for distributing guaranteed bandwidth among service groups in a network node.

DETAILED DESCRIPTION

[0013] According to embodiments described herein, a system and method for distributing guaranteed bandwidth among service groups in a network node are described. The following discussion is presented in the context of an Asynchronous Transfer Mode (ATM) network. It should be understood that the present invention is not limited to ATM networks and may be implemented with other types of networks.

[0014] **Figure 1** is a block diagram of one embodiment for a network. As illustrated in Figure 1, in one embodiment, network 100 includes several network nodes 110 connected through single communication links 120. In one embodiment, network 100 is a data transmission network with guaranteed bandwidth and quality of service requirements, for example an Asynchronous Transfer Mode (ATM) network. Alternatively, network 100 may be another type of data transmission network with guaranteed bandwidth and quality of service requirements.

[0015] In one embodiment, network nodes 110 are located in the middle of the network 100. Alternatively, nodes 110 may also be located at the edges of the network 100. Users 130 access the network 100 and connect to the network nodes 110 via similar communication links 120. In one embodiment, the illustrated communication links 120 support multiple virtual connections.

[0016] **Figure 2** is a block diagram of one embodiment for a network node within network 100. As illustrated in Figure 2, in one embodiment, network

node 110, for example an ATM switching node, receives information, such as data, along multiple input virtual connections 210 within communication link 120. In one embodiment, the information transmitted along virtual connections 210 is incorporated in multiple data units, for example communication cells. In one embodiment, the cells used to transmit information are data cells. Alternatively, transmitted cells may include other types of cells, for example control cells.

[0017] In one embodiment, switching node 110 receives data units along input virtual connections 210 and transfers the data units to multiple ingress line cards 220 described in detail below. Each ingress line card 220 is further coupled to a corresponding egress line card 240 through switch fabric 250. The data units or cells received by egress line cards 240 are then transmitted along multiple output virtual connections 230 to a destination node 110. In one embodiment, multiple output virtual connections 230 are coupled to each egress line card 240 and transmit the cells to destination node 110.

[0018] **Figure 3** is a block diagram of one embodiment for a line card within network node 110. As illustrated in Figure 3, in one embodiment, ingress line cards 220 and egress line cards 240 have similar structures. Although the following discussion is presented in connection with an ingress line card 220, it is to be understood that it also refers to egress line cards 240.

[0019] In one embodiment, ingress line card 220 includes a memory module 310 and a scheduler module 320, coupled to the memory module 310. In one

embodiment, the memory module 310 stores cells received along the multiple virtual connections 210.

[0020] In one embodiment, the scheduler module 320 monitors transmission of data and selects stored cells to be transmitted along the output virtual connections 230. The memory module 310 and the scheduler module 320 will be described in further detail below.

[0021] **Figure 4** is a block diagram of one embodiment for a memory module within the ingress line card 220 or egress line card 240. As illustrated in Figure 4, in one embodiment, memory module 310 includes multiple service groups 420 containing buffers 430 for storing the cells transmitted along the input virtual connections 210. Each buffer 430 within each service group 420 corresponds to one ATM class of service category used in the characterization of each virtual connection 210, for example CBR, VBR, ABR, or UBR.

[0022] Cells arriving along input virtual connections 210 into the network node 110 are identified based on their respective ATM header, which includes the virtual path identifier (VPI) and the virtual channel identifier (VCI). In one embodiment, each cell is classified based on its corresponding ATM header and a corresponding service group 420 is determined. The cell is then stored in a buffer 430 corresponding to the determined service group and class of service category.

[0023] Scheduler module 320 selects service groups 420 and buffers 430 and retrieves cells according to predetermined maximum service rates for each

service group. The service rates are programmed with a predetermined departure parameter, called Inter Cell Gap (ICG) parameter, which specifies the minimum time gap between successive departures of cells.

[0024] In order to transmit the stored cells along the output virtual connections 230, service groups 420 must be selected and cells must depart from the service groups at scheduled times. The scheduling of service groups 420 uses a Theoretical Departure Time (TDT) variable parameter attached to each service group 420. The TDT parameters for service groups 420 are fixed in a time scale and are spaced by their corresponding ICG. The scheduler module 320 uses the ICG to calculate the next TDT parameter of each service group 420 for a future selection. If the TDT parameter of a service group 420 is less than or equal to a current time counter value on the time scale, then the service group 420 is eligible to release cells to be transmitted on an output virtual connection 230.

[0025] The scheduler module 320 also monitors the number of cells available within each service group 420. With each arrival of cells, a cell counter for each service group 420 is incremented. With each selection of a service group 420, after cells depart the service group, the cell counter is decremented. If the TDT parameter for a service group 420 is less than or equal to the current time counter, cells may depart the selected service group 420, provided that the corresponding cell counter is not zero, showing an empty service group 420.

[0026] At a predetermined selection time, more than one service groups 420 may be eligible and thus contending for selection of a service group. Scheduler module 320 selects the service group 420 having the lowest TDT parameter on the time scale among all the eligible service groups. The remaining service groups 420 are subsequently serviced based on their corresponding TDT parameters. After a service group 420 is selected, its corresponding TDT parameter is updated with its specific ICG in preparation for the next selection of a service group to be serviced.

[0027] Cells must depart from buffers 430 within the selected service group 420 at scheduled times. The scheduling of buffers 430 within the selected service group 420 uses a Theoretical Departure Time (TDT) variable buffer parameter attached to each buffer 430 within the corresponding service group 420. The TDT variable buffer parameters for the buffers 430 are fixed in a time scale and are spaced by their corresponding ICG. The scheduler module 320 uses the ICG to calculate the next TDT buffer parameter value of the buffer 430 for a future selection. If the TDT parameter of a buffer 430 is less than or equal to a current time counter value on the time scale, then that buffer 430 is not meeting its minimum bandwidth requirements and becomes eligible to release cells to be transmitted on an output virtual connection 230.

[0028] The scheduler module 320 also monitors the number of cells available within each buffer 430. With each arrival of cells, a cell counter within each buffer 430 is incremented. With each selection of a buffer 430, after cells depart

the selected buffer, the cell counter of the selected buffer 430 is decremented. If the TDT buffer parameter is less than or equal to the current time counter, cells may depart the selected buffer 430, provided that the corresponding cell counter is not zero, showing an empty buffer 430.

[0029] At a predetermined selection time, more than one buffer 430 may be eligible and thus contending for buffer selection. The scheduler module 320 selects the buffer 430 having the lowest TDT buffer parameter on the time scale to be serviced first. The remaining buffers 430 are subsequently serviced based on their corresponding TDT buffer parameters. After the buffer 430 having the lowest TDT buffer parameter is selected, its corresponding TDT buffer parameter is updated with its specific ICG in preparation for the next selection of a buffer 430 to be serviced.

[0030] In one embodiment, in the network switching node 110, quality of service features allow a certain rate of traffic transmitted along links 120 to be guaranteed, for example a certain minimum cell rate is specified for one class of service. On an ingress path into the switch fabric 250 of the switching node 110, the guaranteed traffic through an ingress line card 220 may compete for bandwidth with best effort traffic received through other ingress line cards 220.

[0031] In one embodiment, the switch fabric 250 primarily uses a static method of selection of traffic, for example a round robin method from each ingress line card 220 to each egress line card 240. When the incident traffic destined for a particular egress line card 240 equals or exceeds a physical

transmission rate of the egress line card 240, and the minimum guaranteed rates of one or more ingress line cards 220 exceed what is distributed by the round robin method of allocation performed by the switch fabric 250, an unfair bandwidth distribution may result, i.e. excess bandwidth traffic may be selected by the switch fabric 250 from one ingress line card 220, while minimum guaranteed bandwidth traffic from other one or more ingress line cards 220 is not satisfied.

[0032] When a service group 420 within an ingress line card 220 fails to meet its minimum guaranteed bandwidth, a speed-up signal is asserted for that particular service group 420. By monitoring the speed-up signal, the scheduler module 320 can prioritize the selection of the service groups 420 by dynamically determining the guaranteed bandwidth requirements of the buffers 430 within the memory module 310 and distributing the entire bandwidth among the service groups 420 under speed-up conditions.

[0033] Speed-up is implemented with a speed-up counter per service group 420 and with global set and reset thresholds. The speed-up counter is incremented, decremented, or cleared, depending on the region where selected buffers 430 within a service group 420 are located. If the selected buffer 430 is in the guaranteed bandwidth region, where its TDT buffer parameter is lower than the current time counter value CT, then it must be serviced in order to meet its minimum guaranteed bandwidth requirements. If the selected buffer 430 is in the excess bandwidth region, where the TDT buffer parameter is

greater than the current time counter value CT, then it has met its minimum guaranteed bandwidth and, if serviced, will send excess bandwidth to the destination.

[0034] When a selected buffer 430 within the service group is not meeting its minimum guaranteed bandwidth, the speed-up counter increments for each selection and reaches the set threshold, where the speed-up signal is asserted for the service group 420. When speed-up is asserted, and the selected buffer 430 within the service group has met its minimum guaranteed bandwidth, the speed-up counter decrements for each selection and reaches the reset threshold, where the speed-up signal is de-asserted. The count is updated for every cell that is released from buffer 430.

[0035] In one embodiment, the scheduler module 320 determines the position on a time scale of a buffer 430 within the service group 420 based on its TDT buffer parameter and the current time counter. If the TDT buffer parameter is greater than the current time counter value, then the buffer 430 is in the excess bandwidth region. As a result, the speed-up signal must be deasserted and the speed-up counter must be reset.

[0036] If the TDT buffer parameter is lower than the current time counter value, and if the difference between the current time counter value and the TDT buffer parameter is greater than twice the corresponding ICG of the buffer, the buffer 430 is selected and the speed-up signal is asserted or maintained. At the same time, the speed-up counter is incremented. The TDT buffer parameter of

the selected buffer 430 is subsequently updated with the corresponding ICG value. At this time, the speed-up signal is asserted and priority is given to the buffer 430 to satisfy the guaranteed bandwidth of buffer 430. The procedure is repeated until the difference between the current time counter and the TDT buffer parameter becomes lower than twice the corresponding ICG of the buffer.

[0037] At the same time, the updated TDT buffer parameter is also compared to the current time counter. For each selection, the TDT buffer parameter is updated with the ICG value until the updated TDT parameter becomes higher than the current time counter value. Then, the speed-up counter is reset and the speed-up signal is deasserted. As a result, the bandwidth used by the increased traffic in the buffer 430 is accounted for as guaranteed bandwidth, thereby providing a means for allocating up to 100 percent of guaranteed bandwidth across ingress line cards 220 through the switch fabric 250 to a destination egress line card 240.

[0038] In one embodiment, several service groups 420 may have their respective speed-up signals asserted at the same time. In this embodiment, the scheduler module 320 further prioritizes the service group selection among the several service groups 420 having the speed-up signal asserted in order to ensure that only service groups 420 which are not meeting their respective guaranteed bandwidths are selected for increased traffic. In an alternate embodiment, other eligible service groups 420, but without the speed-up signal

asserted, may compete with service groups having the speed-up signal asserted. In this embodiment, scheduler module 320 selects the service groups with speed-up signal over the service groups without speed-up signal.

[0039] In one embodiment, the speed-up signals from the service groups 420 are passed downstream to destination queues in order to prioritize the traffic across the switch for those service groups 420 demanding more bandwidth to satisfy their guaranteed bandwidth.

[0040] **Figure 5** is a flow diagram of one embodiment for a method for distributing guaranteed bandwidth among service groups in a network node. According to Figure 5, at processing block 510, the position on a time scale of a buffer storing data units is determined. In one embodiment, the TDT parameter of the buffer is ascertained.

[0041] At processing block 520, a decision is made whether the TDT parameter of the buffer is lower or equal to the value CT of the current time counter. If the TDT parameter of the buffer 430 is greater than the current time counter value, then at processing block 525, the speed-up counter is reset and the speed-up signal is deasserted. Next, at processing block 527, the buffer is selected for release of data units. At processing block 529, the TDT parameter of the selected buffer is updated with its corresponding ICG and blocks 510 through 529 are repeated.

[0042] Otherwise, if the TDT parameter of the buffer is lower or equal to the value of the current time counter, then at processing block 530, a decision is

made whether the difference on the time scale between the value CT of the current time counter and the value of the TDT parameter is greater than twice the value of the inter cell gap (ICG).

[0043] If CT – TDT is lower than twice the ICG, then at processing block 535, a decision is made whether the speed-up counter is equal to zero. If the speed-up counter equals zero, then the process jumps to processing block 540. Otherwise, if the speed-up counter is not zero, then the speed-up counter is decremented at processing block 537.

[0044] Next, at processing block 540, after the speed-up counter is decremented, a decision is made whether the speed-up counter is equal to the reset threshold. If the speed-up counter is not equal to the reset threshold, the process jumps to processing block 527, where the buffer is selected for release of data units. Otherwise, if the speed-up counter is equal to the reset threshold, the process jumps to processing block 525, where the speed-up counter is reset and the speed-up signal is deasserted.

[0045] If the difference between the current time counter value CT and the TDT parameter of the buffer is greater than twice the ICG, at processing block 550, a decision is made whether the speed-up counter has reached a predetermined maximum value (MAX VALUE). If the speed-up counter is at MAX VALUE, then the process jumps to processing block 570. Otherwise, if the speed-up counter is below the MAX VALUE, at processing block 560, the speed-up counter is incremented.

[0046] At processing block 570, a decision is made whether the speed-up counter has reached a predetermined set threshold. If the speed-up counter is lower than the set threshold, the process jumps to processing block 527, where the buffer is selected for release of data units. Otherwise, if the speed-up counter is greater or equal to the set threshold, at processing block 580, the speed-up signal and the speed-up counter are asserted or maintained. The process then jumps to processing block 527, where the buffer is selected for release of data units.

[0047] It is to be understood that embodiments of this invention may be used as or to support software programs executed upon some form of processing core (such as the CPU of a computer) or otherwise implemented or realized upon or within a machine or computer readable medium. A machine readable medium includes any mechanism for storing or transmitting information in a form readable by a machine (e.g., a computer). For example, a machine readable medium includes read-only memory (ROM); random access memory (RAM); magnetic disk storage media; optical storage media; flash memory devices; electrical, optical, acoustical or other form of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.); or any other type of media suitable for storing or transmitting information.

[0048] In the foregoing specification, the invention has been described with reference to specific exemplary embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without

departing from the broader spirit and scope of the invention as set forth in the appended claims. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.